Intelligent model for active power prediction of a small wind turbine

FRANCISCO ZAYAS-GATO*, *Department of Industrial Engineering, CITIC; University of A Coruña; CTC, Avda. 19 de Febrero s/n; 15405; Ferrol, A Coruña, Spain.*

ESTEBAN JOVE, *Department of Industrial Engineering, CITIC; University of A Coruña; CTC, Avda. 19 de Febrero s/n; 15405; Ferrol, A Coruña, Spain.*

JOSÉ-LUIS CASTELEIRO-ROCA, *Department of Industrial Engineering, CITIC; University of A Coruña; CTC, Avda. 19 de Febrero s/n; 15405; Ferrol, A Coruña, Spain.*

HÉCTOR QUINTIÁN, *Department of Industrial Engineering, CITIC; University of A Coruña; CTC, Avda. 19 de Febrero s/n; 15405; Ferrol, A Coruña, Spain.*

FRANCISCO JAVIER PÉREZ-CASTELO, *Department of Industrial Engineering, CITIC; University of A Coruña; CTC, Avda. 19 de Febrero s/n; 15405; Ferrol, A Coruña, Spain.*

ANDRÉS PIÑÓN-PAZOS, *Department of Industrial Engineering, CITIC; University of A Coruña; CTC, Avda. 19 de Febrero s/n; 15405; Ferrol, A Coruña, Spain.*

ELENA ARCE, *Department of Industrial Engineering, CITIC; University of A Coruña; CTC, Avda. 19 de Febrero s/n; 15405; Ferrol, A Coruña, Spain.*

JOSÉ LUIS CALVO-ROLLE, *Department of Industrial Engineering, CITIC; University of A Coruña; CTC, Avda. 19 de Febrero s/n; 15405; Ferrol, A Coruña, Spain.*

Abstract

In this study, a hybrid model based on intelligent techniques is developed to predict the active power generated in a bioclimatic house by a low power wind turbine. Contrary to other researches that predict the generated power taking into account the speed and the direction of the wind, the model developed in this paper only uses the speed of the wind, measured mainly in a weather station from the government meteorological agency (MeteoGalicia). The wind speed is measured at different heights, against

*E-mail: f.zayas.gato@udc.es

Vol. 31, No. 4, © The Author(s) 2022. Published by Oxford University Press.

Advance Access published 1 March 2022 https://doi.org/10.1093/jigpal/jzac040

the usual measurements in others researches, which uses the wind speed and the direction measured in a weather station on the wind turbine nacelle. The prediction is performed 30 minutes ahead, what ensures that the Building Management System knows the energy generated by the low power wind turbine 30 minutes before, and it can adapt the consumption of different equipment in the house to optimize the power use. The main objective is to allow the Building Management System to optimize the uses of energy, taking into account the predicted amount of energy that will be produced and the energy consumed in the house. The developed model uses a hybrid topology with four clusters to improve the prediction, achieving an error lower than 6.5% for Mean Absolute Error measured in a final test. To perform this test, part of the original dataset was isolated from the beginning of the training process to check the model with a dataset that is not used before, simulating the model as it is receiving new data.

Keywords: Intelligent model, hybrid model, low power wind turbine, microgrid, power prediction, energy use optimization.

1 Introduction

In recent decades, the introduction of renewable resources for energy generation has become a fundamental pillar of the global energy mix. The wind energy provides a fluctuating and environmentally feasible possibility, in contrast to the different renewable energy technologies available. Moreover, the wind energy could increase the sustainability of the Spanish energy system and reducing the impact of fossil fuels [18]. This fact turns wind energy into a suitable alternative both on a large scale (wind farms) and for small consumers (wind turbines in houses—microgrids). However, compared to traditional technologies characterized by stable and widely studied generation profiles, renewable sources depend strongly on the presence of natural resources. This poses a completely new problem from the point of view of generation; not only does the variability of demand come into play, but also the availability of a resource that varies constantly [44].

This has led to the development of new techniques and studies focused on predicting the evolution of these resources. As opposed to traditional techniques based on meteorological models, artificial intelligence and, specifically, machine learning has allowed for the improvement of predictions to characterize variations in the natural resource. Most of the previously published studies that predict the power generate by renewable resources are focused on the modeling of wind variation [33, 36, 39, 42], solar irradiance [4, 41] or tidal power [6]. The artificial intelligence techniques have been shown to outperform the traditional models. Some other studies have focused on the prediction of energy demand with the aim of optimizing generation strategies and the use of resources [9]. Hydrogen is one of the future energies because it reduces the greenhouse gas emission compare with fossil fuels; there are some researches in this field, like [7], which uses artificial intelligence to predict the necessary hydrogen inlet to produce the desired power. The current trend is to combine the two previous problems, so that it is possible to predict, for a time horizon, the energy generated considering the current state of the renewable resource as well as the current performance of the system. Thus, introducing this information, together with an intelligent consumption pattern, would make possible to optimize the energy generation. This task is especially relevant for small domestic consumers, where it is easier to adapt the consumption curves to the availability of the energy resource.

Against the usual artificial intelligence algorithms, the combination of these techniques in a hybrid intelligent model improves the prediction performance; e.g. [3, 8] develops a hybrid model to predict the output temperature of a thermal solar collector. This hybrid model technique allows the combination of different heterogeneous algorithms, which complement each other to obtain the best prediction, adapting to the characteristics of the dataset [20, 22]. While most of the studies presented so far focus on the predictions of large power generation, a rigorous modeling of small consumer power generation can have a great impact [12, 26] because nearly 40% of the energy consumed in Europe is located at homes and small power installations [23].

Previous studies demonstrated the effectiveness of using Artificial Neural Network (ANN) models to predict system power output based on wind speed [24] and wind power needed to meet the building energy demand [32]. The influence of meteorological variables on the performance of heating or cooling equipment [34], or to identify the optimal building location [13], among others. The main novel aspect of the present study is the combination of predictive models of available wind energy (wind speed), with future facility performance, and the use of a weather station from the government meteorological agency (MeteoGalicia). For this purpose, a well-tested reference weather dataset was used [2]. Renewable energy is a variable energy resource, which is a challenge. In other words, production is fluctuating, so energy needs to be harnessed when it is available. Thus, optimizing the use of energy when it comes from renewable sources is a very important issue. The use of hybrid models simplifies the internal local models. This feature allows the application of the developed model in a low-power installation, because no high performance equipment is needed to run the model.

Therefore, this work is focused on the proposal and analysis of hybrid models for the prediction of power generation from a small wind turbine installed in a house. Specifically, the viability of this method will be analysed on the low power wind turbine installed in the bioclimatic house of the Sotavento Experimental Wind Farm, in the province of Lugo, Spain. This installation is created to study the use of several renewable energies in home. In this research, a hybrid intelligent model to predict the power generated by a wind turbine was created. The main reason to predict directly the power is that this model will include all the losses of the energy system, and the output of the model could be used by the Building Management System to optimize the energy consumed, by adapting the different loads in the house.

The remainder of the paper is organized as follows: Section 2 introduces the methodology, the analysed building, the wind turbine and the dataset. The Model Approach and the procedure to create a hybrid intelligent model are described in Section 3, and also the used algorithms. Subsequently, the configuration and the obtained performance of the model are shown in Section 4. The conclusions of this study are provided in Section 5.

2 Methodology

Figure 1 displays a summary of the methodology used and described in this section. Workflow begins with facility target description. This research is focused on the wind speed forecasting in order to predict the energy generated by the wind turbine installed in the building. Wind turbine is used to provide the electricity demand. The weather data files used for the prediction model were obtained from two sources: the MeteoGalicia historical database, available at the agency webpage, and onsite weather station. A four-stage hybrid prediction model was used to make the wind speed forecasts. Finally, error metrics were used to evaluate the models results.

2.1 Case Study

The target building, shown in Figure 2, was built in 2009. Designed as a single floor dwelling, with a net area of 219 m^2, this house is not used as a residential building, but as a visiting and exhibition space. It is located in Lugo (northwest coast of Spain), facing south-north. The climate

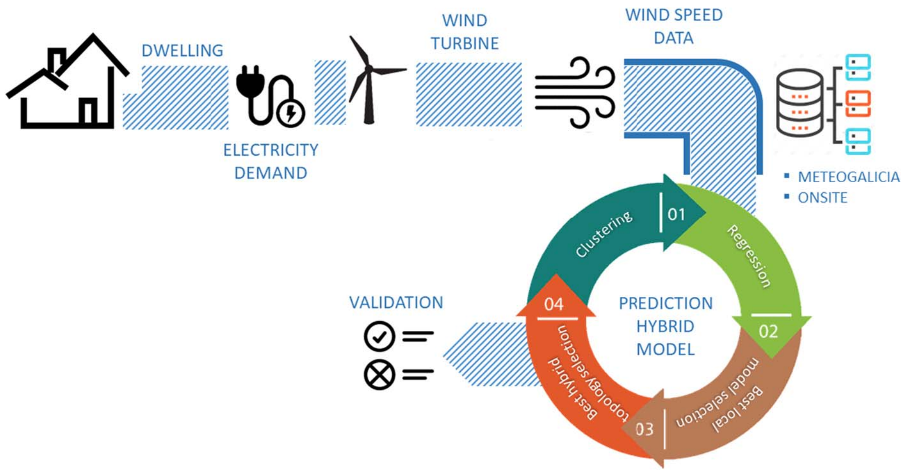

FIGURE 1. Methodology diagram.

FIGURE 2. Sotavento bioclimatic house.

for the area (Lugo), located at a latitude of 43.355062°N and longitude of 7.880769°W, is oceanic [1]. According to the Köppen–Geiger classification, the Cfc marine cool winter and summer-mild with no dry season climate is predominant in the dwelling location.

The structure, material and location of this building were designed to optimize the energy efficiency. The needs of Domestic Hot Water (DHW) and electricity are fulfilled using the technologies shown in Figure 3.

The demand of electricity is covered in three different ways:

1. **Wind turbine.** A low power wind turbine, with eight meters of height, is capable of generating up to 1.5 kW.
2. **Photovoltaic.** Twenty-two panels of polycrystalline silicon material are connected to produce a total power of 2.7 kW. They are placed near the thermal panels with South, East and West orientation.
3. **Power grid.** This part supplies electric energy when the renewable sources cannot cover the demand.

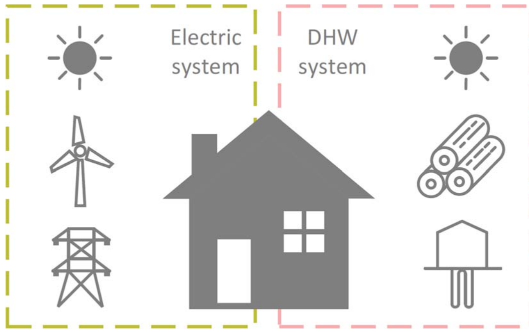

FIGURE 3. Electric and DHW system.

FIGURE 4. Low power wind turbine installed in Sotavento bioclimatic house.

On the other hand, the production of thermal energy for DHW and heating is achieved through the following technologies.

1. **Geothermal.** A horizontal heat exchanger is in charge of extracting the heat from the ground through a five hundred meter pipe.
2. **Solar thermal.** The sun radiation is exploited through eight solar panels placed on the roof. They cover a total surface of 20 m^2, and the solar thermal energy is accumulated using a 1000 l storage tank.
3. **Biomass.** This system is designed to deliver a thermal power that can be regulated from 7 to 20 kW. This is done by means of pellets stored in a silo as biofuel.

2.1.1 Wind turbine The wind is converted to rotatory mechanic movement using a BORNAY 1500 system [1], Figure 4. It has two carbon fibre blades coupled to a three phase Permanent Magnets Synchronous Generator (PMSG) that converts the motion to electric power (1500 W). This generator is located in a housing structure that ensures the proper system orientation.

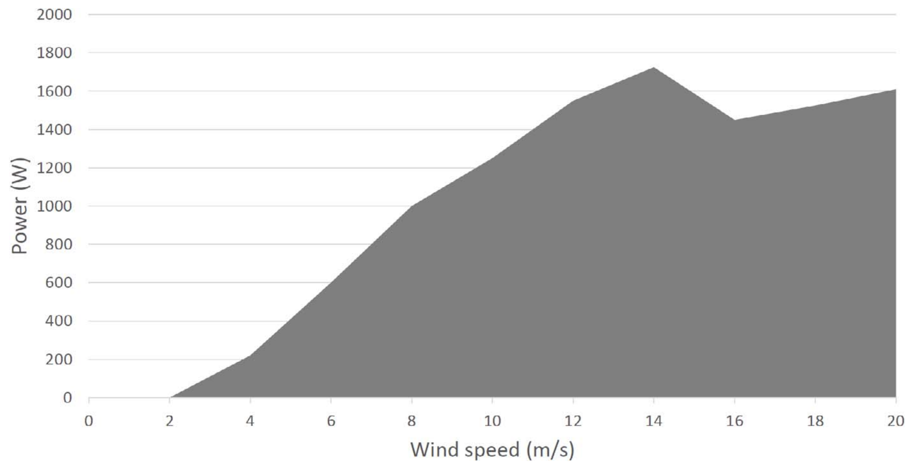

FIGURE 5. Electric power over wind speed. Plot created with the data from the manufacturer [1].

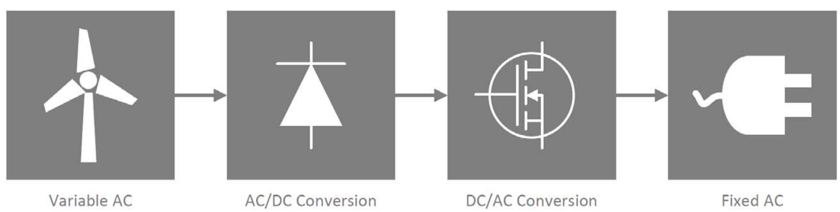

FIGURE 6. Diagram of electric conversion.

The power generated depends on the wind speed as it is shown in Figure 5. In this case, the highest power (1700 W) is achieved with a wind speed of around 14 m/s, and with a wind speed lower than 2 m/s the wind turbine do not produce any power.

The rotational movement of the PMSG produces electric energy in the form of an Alternating Current (AC) whose frequency varies depending on the angular speed. Since voltage waveform in the electric plug must have fixed frequency and fixed peak voltage, the process shown in Figure 6 is carried out. First, a rectifying stage converts the AC energy produced into Direct Current (DC) electrical energy of 48 V by means of semiconductors, capacitors and voltage regulators. This DC value is fixed regardless of the frequency of the AC generated at the PMSG. Then, an inverter transforms the DC into an AC with 220 V and 50 Hz. This is done through a BORNAY AURORA 3600 inverter. In [1] this installation is explained with more details.

2.1.2 Dataset This research has been developed using data collected from April 2017 to March 2018, which represents a whole Typical Meteorological Year and it allows to use the created model to predict the power at any time in a year. They are measured with a sample rate of 10 minutes, resulting in a dataset with 52,560 samples.

Two sources were used to generate the dataset: the nearest weather station of the MeteoGalicia agency network, and the on-site weather station of Sotavento. MeteoGalicia observation network is composed of 166 weather stations distributed throughout Galicia, northwestern Spain. The

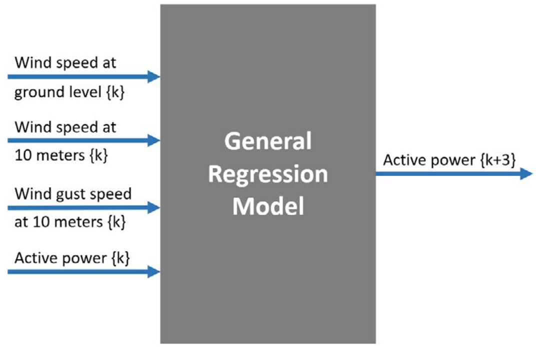

FIGURE 7. Model approach.

specific MeteoGalicia weather station used in this research is ‘Marco da Curra’, located at a latitude of 43.491467°N and longitude of 8.252274°W. Weather station wind speed and wind gust measurements are provided at a height of 10 m. MeteoGalicia webpage [2] provides free access to current conditions and to all historical recorded datasets. The on-site weather station (Sotavento) provided wind speed data at ground level, and the active power data generated by the wind turbine. Sotavento weather station stores all the data generated in the bioclimatic house to allow conduct future researches. Each sample in the dataset contains the following information:

- Wind speed at ground level (m/s).
- Wind speed at a height of 10 meters (m/s).
- Speed of wind gusts at a height of 10 meters (m/s).
- Active power generated (W).

3 Model approach

The basic input/output schematic of the model used in this paper is the one shown in Figure 7. The model predicts, using data from the current instant, the power generated by the wind turbine 3 instants later, 30 minutes. To perform this prediction, the model uses as inputs: the power generated and the other variables about the wind speed described in the previous section, at the specific instant when the model is running.

As this research develops a hybrid model, the general approach shown in Figure 7 is divided into some local models depending on the final hybrid topology. Figure 8 shows the internal layout of a generic hybrid model, where it is possible to see the selector that ‘connects’ the output of the specific local models, chosen according to the inputs, to the output of the hybrid model.

To create a hybrid model like this, it is necessary to follow the flowchart shown in Figure 9.

1. Firstly, in the clustering phase, the whole dataset used to create the model is divided into several clusters. Usually, the optimal number of clusters is not known previously, and then, some hybrid topologies are created dividing the dataset into different number of clusters.
2. Once the clusters have been created, in the regression phase, several regression algorithms are used to obtain several local models for the same cluster data.

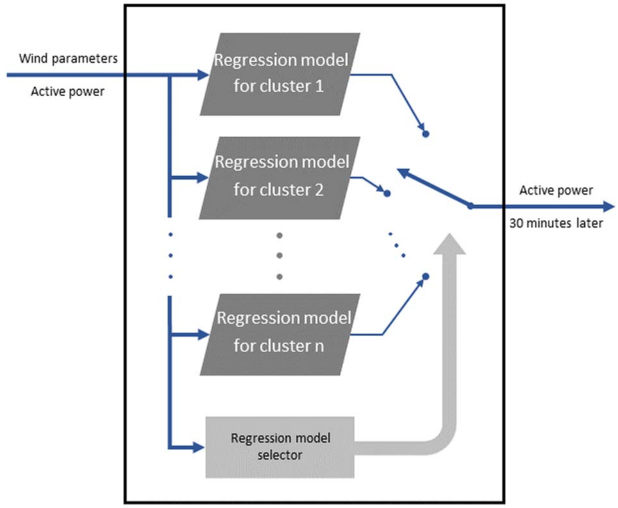

FIGURE 8. Internal diagram to achieve the hybrid model.

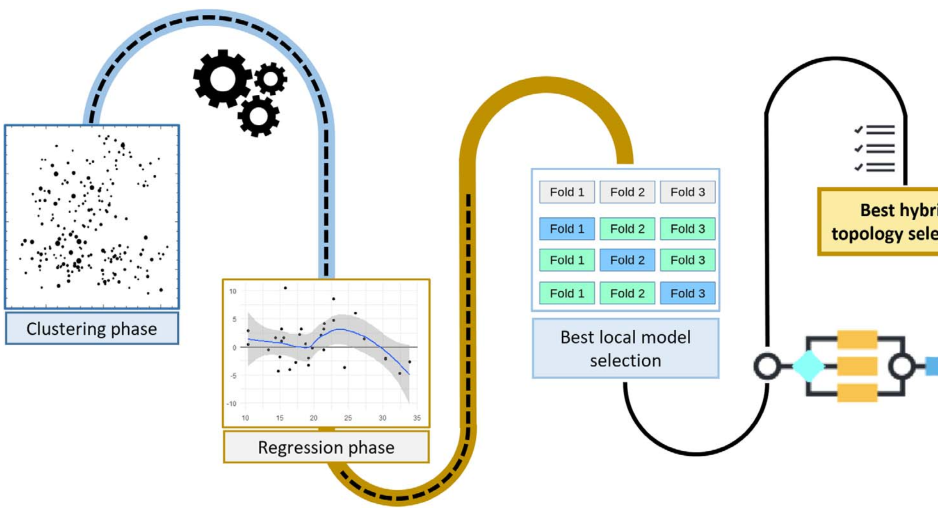

FIGURE 9. Flowchart of the hybrid model creation phases.

3. The best local model is selected depending on the performance of the different regression algorithms used. If the same algorithm has different internal configuration, they are taking into account as different regression algorithm to compare their performance.
4. Finally, it is necessary to choose the best hybrid topology, the number of local models in the final hybrid model. This selection should not only take into account the performance of the local model; it is possible to use a different dataset (that was not used in regression phase), or to calculate a pondered error based on the number of samples on each cluster.

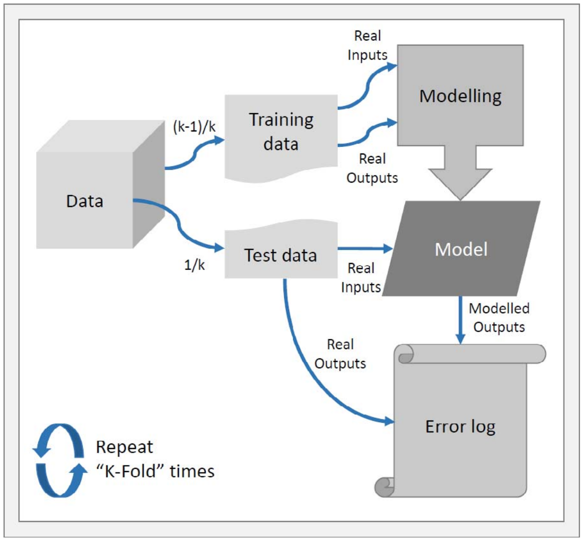

FIGURE 10. K-Fold training and test data selection.

To calculate the performance of each regression algorithm it could be used a Hold Out validation, with a fixed division of the dataset in training and testing data; or a K-Fold cross validation. The last one is the validation technique used in this paper because it produce a more realistic error measurement as it uses all the available data. This type of validation needs to create k models for the same regression algorithm before the error can be calculated. Figure 10 shows a diagram of this validation procedure. The data (for each cluster) are divided k times: the training data are created by adding *k-1* subsets, and the test data use the other subset. The model is training with the corresponding dataset, and then it is tested with the test dataset, storing the real and the predicted output variable. This procedure is repeated until all the subsets have been used as test data, then the error value can be calculated by using all the samples of the cluster. The third phase of the Figure 9, the selection of the best local model for each cluster, is performed by using the error calculated with K-Fold cross validation for all the regression algorithms, and also for the different internal configurations.

As the local models created previously do not use all the data in the clusters, as only $\frac{k-1}{k}$ samples were used for training, it is necessary to train again all the models once the regression algorithms are chosen. It must be highlighted that there is not only one model with the selected regression algorithm, conversely, one local model should be trained (for every cluster) but now using all data available in each cluster.

The final phase to create the hybrid model is the selection of the hybrid topology. In this research a new dataset, isolated from the training phase, was used to test each hybrid topology. To ensure that all the clusters were tested, this dataset was created with data from all the clusters. The best hybrid topology will be the one that achieves the lowest error using this dataset.

3.1 K-Means algorithm

To create the aforementioned clusters, the K-Means algorithm has been used. This is a well-known clustering algorithm that is used to divide a dataset into K different clusters. The algorithm tries to associate samples with similar characteristics [19, 28, 30], and each cluster is defined by a centroid.

The training of the algorithm starts with the selection of the initial centroids, usually randomly chosen. Then, each sample is assigned to a clusters according to the distance between the samples and the centroids; the sample will belong to the cluster defined by the closest centroid. New centroids are calculated as the center of the cluster by using the previous samples assignation. The training ends when the centroids are the same over two iterations [21, 37].

The final centroids are used to assign new samples to their centroids. Euclidean distance is usually used, and the assignation of new samples is very fast [31].

3.2 Artificial Neural Networks

The ANN is one of the best known algorithms in Artificial Intelligence; it can be used for regression and for classification. An ANN is created by the union of several neurons that link the inputs to the output of the net. They usually obtain very good results when they try to predict cases that have not been trained before [14, 15].

The neurons are organized in layers; all the neurons belonging to a layer have the same inputs and outputs. The input and output layer connect the net with the model input and output variables. Moreover, in a feedforward ANN, the internal or hidden layers connect the outputs from the previous layer to the next layer [16]. Each neuron has weights for all the inputs and a *bias*, which is a constant input. All the neuron inputs are pondered by their weights, and then they are added to obtain a unique value. This value is used as the input of the internal activation function for each neuron, and the output of the neuron is the output of that function [27]. This output defines the state of the neuron, normally in a range from 0 to 1, or -1 to 1 [11].

The Multi-Layer Perceptron is the ANN topology used in this research, it is a feedforward network configured with only one hidden layer and uses the same activation function for all the neurons in each layer. The activation functions can be step, log-sigmoid, tan-sigmoid and linear. As they are used for regression, the configuration for this paper uses the tan-sigmoid function in the internal layer, and the linear function in the output one.

3.3 Polynomial regression

Polynomial regression is an algorithm that achieves a model using a linear function. It is an old technique based on the summation of several basis functions [5, 17, 40, 43]. The number of these basis functions depends on the degree of the internal polynomial and the number of inputs. As an example, Equations (1) and (2) show the model for first and second degree for two inputs: x_1 and x_2.

$$F(x) = b_0 + b_1x_1 + b_2x_2 \tag{1}$$

$$F(x) = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + b_4x_1^2 + b_5x_2^2 \tag{2}$$

The training for these types of algorithms adjusts the internal parameters of the model; in Equation (1) and (2) they are shown as b_0, b_1, b_2, b_3, b_4 and b_5.

3.4 Support Vector Machines for regression

Support Vector Machines is an algorithm used initially only for classification, but there is a modification of this technique that can be used for regression. Basically, this algorithm is based on a transformation of the data to increase the dimension of its representation. In the regression case, once the data are transformed, a linear regression is performed with the new high-dimensional representation.

This research uses a modification of the algorithm that only needs to tune two parameters: γ, the weight vector, and σ, the kernel width. This modified algorithm is called Least Square-Support Vector Regression (LS-SVR) [35, 38].

3.5 Data processing

The data used in this research, as is explained, were acquired with a sample time of 10 minutes between each sample. The data include several wind speeds and the power generated by the low power wind turbine.

The initial dataset shows a great number of bad measurements, and it was necessary to perform an analysis to discard these bad measurements. Moreover, it is necessary to have continuous samples because the developed model uses a predicted horizon of 3 instants. This process in the dataset finally divides the data into 209 subsets with chronological data without failures between samples. After that, the new dataset was created taking into account that the model will predict the generated power 30 minutes ahead (3 instants with a ten minutes sample time). Also, the samples with no power generation were removed. After this analysis, the final dataset had 33,542 samples.

A validation dataset has been extracted from this new dataset, samples that will be used to test the hybrid model once it is created (1,677 samples). The rest of the data, 31,865 samples, was normalized before it was used in the hybrid model creation.

To select the best hybrid topology (fourth step in Figure 9), after creating the clusters data and before start the local models training procedure (between first and second step in Figure 9), 5% of the samples for each cluster were isolated. This dataset is not be included in the training process.

3.6 Performance measurements

The performance of the different regression models created in this research is measured with an error value. Different error values were calculated taking into account the desired value of the output (y_i) and the predicted value ($\hat{y}_i$), the output calculated by the model. In the specific case of this paper, the performance is calculated between the real power and the output of the different models.

The Mean Squared Error (MSE) is one of the most widely used errors to measure the performance of a regression model. Equation 3 shows the formula used to calculate this value, where n is the number of samples.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \tag{3}$$

The Mean Absolute Error (MAE) is another commonly used error measurement; specifically, in this research, the MSE and MAE are used together to choose the best regression algorithm. Equation 4 shows the formula to calculate this error.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid \hat{y}_i - y_i \mid \tag{4}$$

The Normalized Mean Squared Error (NMSE) is used to measure the error that takes into account the variance of the values. This measure is used, in this research, to calculate the performance with the validation dataset, with the achieved model. Equation 5 shows the formula used.

$$NMSE = \frac{MSE}{\sigma^2} \tag{5}$$

Variables in the equations:
n - Number of samples to calculate the error value
$\hat{y}_i$ - Predicted value, the output of the model
y_i - Desired value
σ - Variance of error distribution

4 Results

In this research, the results were divided into four different subsections following the procedure of creating a hybrid model; moreover, a validation test of the approach is included. It must be highlighted that three completely different datasets were created:

1. The first one is a dataset to test the final hybrid model, and it is randomly chosen from the initial data without taking into account the clusters.
2. The second one is a dataset with the 5% of data from each cluster, and it will be used to select the best hybrid topology.
3. The third one contains the data to train the local models.

4.1 Clustering results

The K-Means algorithm creates as many clusters as is required by the user. In this paper, the dataset was divided up to nine times to create 9 subsets of data. The number of clusters was varied from 2 to 10 to obtain 10 different sets of data to train the models: the global model (with only one global *cluster*) and 9 hybrid topologies.

The K-Means was trained with random initial centroids and, to avoid local minima during the training, the procedure to calculate the final centroids was repeated 20 times. The number of samples in each of the clusters is shown in Table 1 where each column represents the hybrid topology configuration for each model, and the rows the different clusters inside the hybrid models. This way to show the results in a matrix will be the same for the other results sections where local model parameters are shown. In this case, the number of samples, it can be seen that there are some clusters with less samples than others; in this research, the clustering produced unbalanced groups, but with enough data to generalize the obtained results.

4.2 Modeling results

K-Fold cross validation was used to select the best regression algorithm for each cluster. This validation technique divides the dataset k times and uses a dataset created by join *k-1* subset for training, and the other subset for testing. This procedure was repeated k times until all the subsets were used for testing. With K-Fold cross validation, the error measurement takes into account all the datasets, and the obtained value is more realistic than if Hold Out validation were used.

TABLE 1. No of samples for each subset.

	Global	Hyb. 2	Hyb. 3	Hyb. 4	Hyb. 5	Hyb. 6	Hyb. 7	Hyb. 8	Hyb. 9	Hyb. 10
C-1	30,272	7,850	3,694	1,916	1,282	1,219	1,129	780	546	485
C-2		22,422	11,286	5,749	3,644	1,621	1,148	1,047	727	703
C-3			15,293	10,787	7,218	3,291	3,258	2,276	1,769	1,719
C-4				11,821	7,541	6,382	4,124	3,670	2,140	2,004
C-5					10,587	7,436	4,423	4,367	3,433	2,008
C-6						10,324	6,589	4,403	3,677	3,196
C-7							9,603	5,280	4,387	3,705
C-8								8,550	5,238	4,446
C-9									8,356	4,525
C-10										7,482

TABLE 2. Mean Square Error for each individual hybrid model.

	Global	Hyb. 2	Hyb. 3	Hyb. 4	Hyb. 5	Hyb. 6	Hyb. 7	Hyb. 8	Hyb. 9	Hyb. 10
C-1	**0.0031**	0.0077	0.0111	0.0146	0.0174	0.0160	0.0124	0.0185	0.0175	0.0179
C-2		**0.0015**	0.0032	0.0053	0.0070	0.0100	0.0163	0.0128	0.0180	0.0175
C-3			**0.0011**	**9.9302e^{-4}**	0.0033	0.0065	0.0064	0.0083	0.0065	0.0066
C-4				0.0022	**8.9082e^{-4}**	0.0024	0.0030	0.0041	0.0079	0.0080
C-5					0.0016	**8.7083e^{-4}**	0.0021	0.0029	0.0023	0.0058
C-6						0.0015	**8.5822e^{-4}**	0.0015	0.0042	0.0028
C-7							0.0013	**8.3187e^{-4}**	0.0015	0.0021
C-8								0.0012	**8.1921e^{-4}**	0.0012
C-9									0.0011	**7.9505e^{-4}**
C-10										0.0011

All the regression algorithms used in this research have been tested with K-Fold, and, if there were internal parameters to adjust, as many models were created as configurations to test.

- The ANNs have been configured with one hidden layer, using tan-sigmoid as activation function for the internal neurons (in the hidden layer), and the linear function in the neuron of the output layer. The number of internal neurons was varied from 1 to 15, and the internal weights were adjusted using Levenberg–Marquardt as the optimization algorithm. There were 15 ANN models tested for each local model.
- The Polynomial Regression has been configured to test the first and second degree. There were 2 Polynomial Regression models tested for each cluster.
- The LS-SVR has an internal function to auto-tune the internal parameters, for that reason, it has only tested one LS-SVR model for each local model.

Tables 2 and 3 show the error for each local model. The first table shows the lowest MSE obtained for each cluster, and the second one shows the MAE. Table 4 shows the best regression algorithm for each cluster, which, in this research, is based on the lowest MSE value.

It is possible to see that the lowest MSE is achieved in the 9th cluster of the hybrid topology that divides the dataset 10 times, while the highest MSE was in the 1st cluster of the hybrid topology with 8 groups. In the case of MAE best and worst error are achieved in the same clusters tan MSE.

TABLE 3. Mean Absolute Error for each individual hybrid model.

	Global	Hyb. 2	Hyb. 3	Hyb. 4	Hyb. 5	Hyb. 6	Hyb. 7	Hyb. 8	Hyb. 9	Hyb. 10
C-1	**0.0365**	0.0637	0.0792	0.0938	0.1029	0.0989	0.0862	0.1065	0.1029	0.1039
C-2		**0.0268**	0.0401	0.0530	0.0632	0.0760	0.0992	0.0878	0.1061	0.1042
C-3			**0.0235**	**0.0218**	0.0408	0.0608	0.0603	0.0693	0.0612	0.0616
C-4				0.0327	**0.0207**	0.0355	0.0389	0.0465	0.0675	0.0677
C-5					0.0280	**0.0207**	0.0334	0.0385	0.0333	0.0564
C-6						0.0270	**0.0204**	0.0279	0.0472	0.0381
C-7							0.0260	**0.0201**	0.0282	0.0318
C-8								0.0248	**0.0199**	0.0255
C-9									0.0238	**0.0197**
C-10										0.0236

TABLE 4. Configuration for each individual hybrid model.

	Global	Hyb. 2	Hyb. 3	Hyb. 4	Hyb. 5	Hyb. 6	Hyb. 7	Hyb. 8	Hyb. 9	Hyb. 10
C-1	ANN5	ANN7	Poly1	ANN4	Poly1	Poly1	ANN1	Poly1	ANN1	ANN2
C-2		ANN5	ANN2	ANN1	ANN1	ANN1	Poly1	ANN1	ANN1	ANN2
C-3			ANN1	ANN2	ANN1	ANN2	ANN1	ANN1	ANN1	ANN1
C-4				ANN3	Poly2	ANN2	ANN1	ANN1	Poly1	ANN2
C-5					ANN1	ANN3	ANN4	ANN1	ANN1	ANN1
C-6						ANN2	Poly2	ANN4	ANN3	ANN2
C-7							ANN2	Poly2	ANN3	ANN1
C-8								ANN2	ANN2	ANN3
C-9									ANN2	ANN1
C-10										ANN3

4.3 Hybrid topology selection

Once the best algorithm for each cluster is selected, the local models have to be trained using all the dataset available in each cluster. As it was explained, the selection of the regression algorithm uses K-Fold cross validation, and then, not all the data were used previously to create the models.

To choose the hybrid topology, the aforementioned dataset has been used, created with 5% of the cluster data. This dataset has been isolated before the training phase, the creation of the local models. These data were used as new samples in the hybrid models, using the centroids created with K-Means algorithm to assign the correct cluster to each sample, and calculate the predicted output with the regression local model.

Table 5 shows the result for this test. The table includes all the hybrid configuration, and also the global model, and the MSE and the MAE for the test dataset are shown. In this table, the error shown is the hybrid topology error, not the local model error; the samples were assigned to each cluster, and the different predicted outputs were calculated with the corresponding internal local model.

The best results have been obtained with a hybrid model with 4 clusters, which achieve an MSE of 0.0030192786309358934, and a MAE of 0.036135910080742334 69. This hybrid model includes four local models with ANN with one, two, three and four neurons for the specific clusters.

TABLE 5. Mean squared error for each model.

	Global	Hybrid Model (Local Models)								
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
MSE	0.0031	0.0030	0.0031	**0.0030**	0.0031	0.0031	0.0031	0.0031	0.0031	0.0030
MAE	0.0364	0.0361	0.0364	**0.0361**	0.0364	0.0364	0.0363	0.0362	0.0362	0.0361

TABLE 6. Performance values for the best hybrid configuration.

	MSE	**MAE**
4 local models configuration	$1.261e^{+4}$	73.822
(with normalized values)	0.0033	0.0376

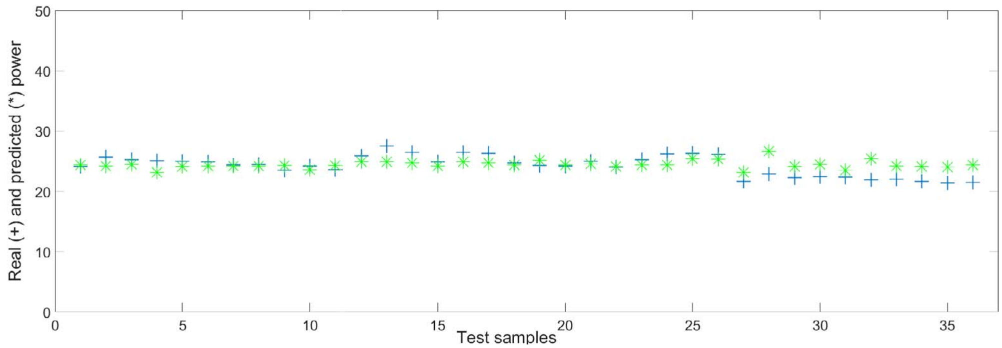

FIGURE 11. Test results.

4.4 Test results

To validate the hybrid model developed in this research, a validation dataset has been used. This dataset was isolated from the original data, and it has never been used in the clusters creation phase, in the training phase or in the hybrid topology selection. It has 1,677 samples randomly chosen from the original dataset acquired in the low power wind turbine described in the case study.

Table 6 shows the error measurements obtained in this final check. It includes the MSE and MAE with real units and also with normalized values. The maximum value of the predicted variable in the validation dataset is 1202.17 W, and it means that the error obtained was less than 6.5%. The NMSE obtained for this test dataset was 0.5208. Moreover, Figure 11 shows the predicted (green asterisk) and the real (blue cross) power generated by the small wind turbine.

These results can be compared with the ones obtained in other papers like [29], where a model is created to predict the power generated by four wind turbines of 1.7 MW. In that paper, the authors predict the power 10 minutes sooner, and the Root Mean Squared Error obtained is between 73.31 and 89.94; in our model, that error achieves a value of 112.29, but the prediction is made 30 minutes ahead.

Other research, like [25], created the model for whole wind turbine, including the internal components. In that paper, the predicted output has a MAE of $5.59e^{+4}$ W, while in ours the MAE is 73.822 W. Most of the previous literature that model the power generated by wind turbines are based on big ones, not on small wind turbines to install in a microgrid environment. For example, in [10], the authors used machine learning to model a 1.5 MW wind turbine, achieving MAE values of around 50 kW in a prediction with a horizon of 10 minutes.

A time horizon of 30 minutes is used in this research. The use of a hybrid topology allows the model to be run on a low-performance computer. That is, it does not require a high-level computer. The estimated error (MAE under 6.5%) is within the range of domestic applications.

5 Conclusions

This paper presents a hybrid intelligent model created to predict the power generated by a small wind turbine. Other studies are centred on wind speed prediction, or in standard wind turbines; the objective of this research are the small ones because they are the ones that can be used in microgrids. Moreover, the importance of predicting power instead of wind is that this value can be used directly by the Building Management System to optimize the energy consumption in a house.

The hybrid model uses wind speed measurements, from weather stations, and the generated power to predict the power generation 30 minutes earlier. The use of a weather station from the autonomous weather agency is another important improvement, because the model does not require that the house, where the small wind turbine is placed, has its own weather station. The final model uses four ANNs as local models, each one with its own internal parameters. It obtains an NMSE of 0.5208, and a MAE lower than 6.5%, with the validation dataset.

It could be mentioned, as a possible future works, the possibility of increasing input variables to include more meteorology parameters like humidity or temperature, to study their possible effect on the generated power. Moreover, it is possible to increase the prediction horizon, to calculate the desired variable earlier.

Funding

Funding for open access charge: University of A Coruña / CISUG.

Acknowledgements

CITIC, as a Research Center of the University System of Galicia, is funded by the Consellería de Educación, Universidade e Formación Profesional of the Xunta de Galicia through the European Regional Development Fund (ERDF) and the Secretaría Xeral de Universidades (Ref. ED431G 2019/01).

Conflict of interest

The authors declare no potential conflict of interests.

References

[1] SPIText *Sotavento installation*. http://www.sotaventogalicia.com, 2020. Accessed: 2020-10-19.
[2] SPIText *Meteogalicia. Observation. Meteorological network*. https://www.meteogalicia.gal, 2021. Accessed: 2021-02-08.

[3] H. Aláiz-Moretón, M. Castejón-Limas, J.-L. Casteleiro-Roca, E. Jove, L. F. Robles and J. L. Calvo-Rolle. A fault detection system for a geothermal heat exchanger sensor based on intelligent techniques. *Sensors*, **19**, 2740, 2019.

[4] J. M. Barrera, A. Reina, A. Maté and J. C. Trujillo. Solar energy prediction model based on artificial neural networks and open data. *Sustainability*, **12**, 6915, 2020.

[5] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer New York, Inc, Secaucus, NJ, USA, 2006.

[6] S. Buhan, D. Kucuk, M. S. Cinar, U. Guvengir, T. Demirci, Y. Yilmaz, F. Malkoc, E. Eminoglu and M. U. Yildirim. A scalable river flow forecast and basin optimization system for hydropower plants. *IEEE Transactions on Sustainable Energy*, **11**, 2220–2229, 2020.

[7] J.-L. Casteleiro-Roca, A. J. Barragán, F. Segura, J. L. Calvo-Rolle and J. M. Andújar. Intelligent hybrid system for the prediction of the voltage-current characteristic curve of a hydrogen-based fuel cell. *Revista Iberoamericana de Automática e Informática Industrial*, **16**, 492–501, 2019.

[8] J.-L. Casteleiro-Roca, P. Chamoso, E. Jove, A. González-Briones, H. Quintián, M.-I. Fernández-Ibáñez, R. A. V. Vega, A.-J. P. Pazos, J. A. L. Vázquez, S. Torres-Álvarez, T. Pinto and J. L. Calvo-Rolle. Solar thermal collector output temperature prediction by hybrid intelligent model for smartgrid and smartbuildings applications and optimization. *Applied Sciences*, **10**, 4644, 2020.

[9] J.-L. Casteleiro-Roca, J. Gómez-González, J. Calvo-Rolle, E. Jove, H. Quintián, B. G. Diaz and J. M. Perez. Short-term energy demand forecast in hotels using hybrid intelligent modeling. *Sensors*, **19**, 2485, 2019.

[10] A. Clifton, L. Kilcher, J. K. Lundquist and P. Fleming. Using machine learning to predict wind turbine power output. *Environmental Research Letters*, **8**, 024009, 2013.

[11] B. M. del Brío and A. S. Molina. *Redes Neuronales y Sistemas Borrosos*. Ra-Ma, 2006.

[12] B. Dong, Z. Li, S. M. M. Rahman and R. Vega. A hybrid model approach for forecasting future residential electricity consumption. *Energy and Buildings*, **117**, 341–351, 2016.

[13] H. Eliasstam, K. N. Genikomsakis and C. S. Ioakimidis. Wind power and electricity consumption forecasting on a smart house location. In *Int. Conf. Renewable Energies and Power Quality (ICREPQ'13)*, 2013.

[14] M. I. A. Galipienso, M. A. C. Quevedo, O. C. Pardo, F. E. Ruiz and M. A. L. Ortega. *Inteligencia Artificial. Modelos, Técnicas y Áreas de Aplicación*. Editorial. Paraninfo, 2003.

[15] J. R. H. González and V. J. M. Hernando. *Redes Neuronales Artificiales: Fundamentos, Modelos y Aplicaciones*. RA-MA, 2000.

[16] A. J. Maren, C. T. Harston and R. M. Pap. *Handbook of Neural Computing Applications*. Elsevier Science, 2014.

[17] R. M. Heiberger and E. Neuwirth. Polynomial regression. In *R Through Excel, Use R*, pp. 269–284. Springer, New York, 2009.

[18] G. M. Joselin Herbert, S. Iniyan, E. Sreevalsan and S. Rajapandian. A review of wind energy technologies. *Renewable and Sustainable Energy Reviews*, **11**, 1117–1145, 2007.

[19] E. Jove, P. Blanco-Rodríguez, J. L. Casteleiro-Roca, J. Moreno-Arboleda, J. A. López-Vázquez, F. J. de Cos Juez and J. L. Calvo-Rolle. Attempts prediction by missing data imputation in engineering degree. In *International Joint Conference SOCO'17-CISIS'17-ICEUTE'17 León, Spain, September 6–8, 2017, Proceeding*, pp. 167–176. Springer, 2017.

[20] E. Jove, J.-L. Casteleiro-Roca, H. Quintiá, J. A. Méndez-Pérez and J. L. Calvo-Rolle. Anomaly detection based on intelligent techniques over a bicomponent production plant used on

wind generator blades manufacturing. *Revista Iberoamericana de Automática e Informática Industrial*, **17**, 84–93, 2020.

[21] E. Jove, J.-L. Casteleiro-Roca, H. Quintián, J.-A. Méndez-Pérez and J. L. Calvo-Rolle. Virtual sensor for fault detection, isolation and data recovery for bicomponent mixing machine monitoring. *Informatica*, **30**, 671–687, 2019.

[22] E. Jove, J.-L. Casteleiro-Roca, H. Quintián, D. Simić, J.-A. Méndez-Pérez and J. L. Calvo-Rolle. Anomaly detection based on one-class intelligent techniques over a control level plant. *Logic Journal of the IGPL*, **28**, 502–518, 2020.

[23] S. Karatasou, M. Laskari and M. Santamouris. Determinants of high electricity use and high energy consumption for space and water heating in european social housing: socio-demographic and building characteristics. *Energy and Buildings*, **170**, 107–114, 2018.

[24] A. Khandakar, M. E. H. Chowdhury, M. K. Kazi, K. Benhmed, F. Touati, M. Al-Hitmi and S. P. G. Antonio Jr.. Machine learning based photovoltaics (PV) power prediction using different environmental parameters of qatar. *Energies*, **12**, 2782, 2019.

[25] A. Kusiak and W. Li. Virtual models for prediction of wind turbine parameters. *IEEE Transactions on Energy Conversion*, **25**, 245–252, 2010.

[26] K. S. Lee, S. Andrew Putrayudha, E. Evgueniy, Y. Libing, E. C. Kang and E. J. Lee. An impact of artificial intelligence control on photovoltaic/thermal (pvt)—ground source heat pump (gshp) hybrid system. In *ISES Solar World Congress 2015, Conference Proceedings*, pp. 1611–1619, 2015.

[27] R. F. López and J. M. F. Fernández. *Las Redes Neuronales Artificiales*. Netbiblo, 2008.

[28] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pp. 281–297, 1967.

[29] M. Mehrjoo, M. J. Jozani and M. Pawlak. Wind turbine power curve modeling for reliable power prediction using monotonic regression. *Renewable Energy*, **147**, 214–222, 2020.

[30] J. Moody and C. J. Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation*, **1**, 281, 6–294, 1989.

[31] J. H. Orallo, M. J. R. Quintana and C. F. Ramírez. *Introducción a la Minería de Datos*. Editorial Alhambra S.A, 2004.

[32] S. G. Rahul, P. Kavitha and G. Dhivyasri. Prediction of electricity load using artificial neural network for technology tower block of vit university. *International Journal of Applied Engineering Research*, **12**, 7717–7723, 2017.

[33] R. Sharma, T. Shikhola and J. K. Kohli. Modified fuzzy q-learning based wind speed prediction. *Journal of Wind Engineering and Industrial Aerodynamics*, **206**, 104361, 2020.

[34] S. Sholahudin and H. Han. Simplified dynamic neural network model to predict heating load of a building using taguchi method. *Energy*, **115**, 1672–1678, 2016.

[35] I. Steinwart and A. Christmann. *Support Vector Machines*. Springer Publishing Company, Incorporated, 2008.

[36] M. Tomás-Rodríguez and M. Santos. Modelling and control of floating offshore wind turbines. *Revista Iberoamericana de Automática e Informática Industrial*, **16**, 381–390, 2019.

[37] P. I. Viñuela and I. M. G. León. *Redes de Neuronas Artificiales: Un Enfoque Práctico*. Pearson Educación—Prentice Hall, 2004.

[38] L. Wang and J. Wu. Neural network ensemble model using PPR and LS-SVR for stock et eorecasting. In *Advanced Intelligent Computing*, pp. 1–8, 2012.

[39] S. Wang, N. Zhang, W. Lei and Y. Wang. Wind speed forecasting based on the hybrid ensemble

empirical mode decomposition and ga-bp neural network method. *Renewable Energy*, **94**, 629–636, 2016.

[40] X. Wu. *Optimal designs for segmented polynomial regression models and web-based implementation of optimal design software*. State University of New York at Stony Brook, Stony Brook, NY, USA, 2007.

[41] W. Zhou, Q. Li and X. Xia. Multi-timescale forecast of solar irradiance based on multi-task learning and echo state network approaches. *IEEE Transactions on Industrial Informatics*, **17**, 300–310, 2021.

[42] X. Yuanyuan and G. Yang. A short-term wind speed forecasting hybrid model based on empirical mode decomposition and multiple kernel learning. *Complexity*, **2020**, 1–13, 2020.

[43] Z. G. Zhang and S. C. Chan. On kernel selection of multivariate local polynomial modelling and its application to image smoothing and reconstruction. *Journal of Signal Processing Systems*, **64**, 361–374, 2011.

[44] A. Zidan and E. F. El-Saadany. Distribution system reconfiguration for energy loss reduction considering the variability of load and local renewable generation. *Energy*, **59**, 698–707, 2013.

Received 20 February 2021